ACHIEVEMENT STANDARD 90647

MODELLING

NOTES FOR TEACHERS

This document has been prepared to provide some background information relevant to, and to assist teachers interpret the requirements of, the achievement standard 90647

ACHIEVEMENT STANDARD 90647 MODELLING

NOTES FOR TEACHERS

This document has been prepared to provide some background information relevant to, and to assist teachers interpret the requirements of, the achievement standard 90647 *Use a mathematical model involving curve fitting to solve a problem*.

GENERAL

This standard is not just about the mechanical process of transforming data to linearise it and attempting to fit different types of model – it is about modelling (one aspect of which is fitting a function to model data). Before embarking on any analysis, students need to obtain an understanding of the variables and not just treat analysis related to this standard as a mathematical exercise. They should think about the data, how it was collected and what the purpose of collecting the data may have been. Gaining an understanding of these aspects is helped by trying to visualise the data collection process.

Students need to show an understanding of the context in which the values obtained have been calculated. This includes giving units where relevant, and rounding to an appropriate precision (which could result in an integral value).

If any of the possible aspects for Excellence that are listed in the standard are not relevant to a particular situation, reference to them should be omitted from the assessment task.

Comments made by students for Excellence need to be justified – e.g. if any other model is proposed to fit the data then the model needs to be realistic and an explanation needs to be given why the proposed model was selected. Comments should also be made in context, i.e. they should not be in general terms but should relate to the specific variables being investigated.

The following headings relate to particular specifications of the achievement standard for Merit and Excellence.

MERIT

1 Selecting a model to test

Evidence of model selection is required. Most commonly this comes from at least one of:

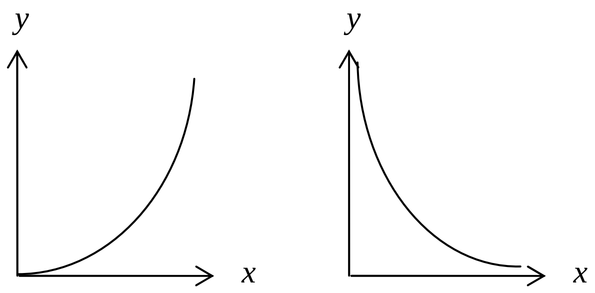
- considering the nature of the underlying variables (see further comments following)
- a visual inspection of different models fitted to the data (showing the two variables with the regression line drawn on each)
- linearising the data and noting the appearance of the log/log or log/linear graph.

Considering the nature of the variables being investigated should be the first step in determining an appropriate model. The other possibilities should only be used if there is little or nothing that the nature of the variables suggests about the form of the model.

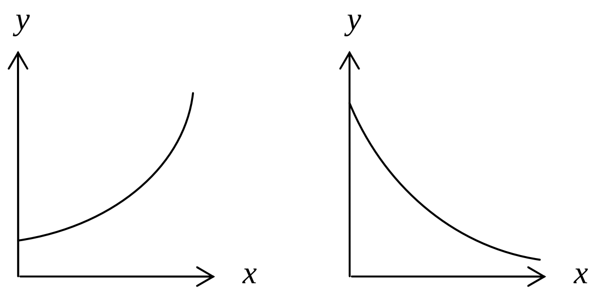
The nature of the variables may suggest the type of model. For example, if the variables are length *l* centimetres of a species and mass *m* grams of the same species, then if it assumed that the species grows uniformly in each of its three dimensions then a model of the form $m = Al^3$ would be appropriate, so that a power law of the form $m = Al^n$ could be tested.

In cases involving growth or decay (decline), an exponential model will be most appropriate. Natural growth or decay phenomena are those where a variable is assumed to be changing at a rate that is proportional to the value of the variable at a given instant, ie $\frac{dy}{dx} \propto y$.

A power law model may be suggested by considerations such as the following. Consider two variables *x* and *y*. Suppose that $x = 0$ means that *y* must be 0, or if $x = 0$ means that *y* cannot exist and values of *x* close to 0 mean that *y* is very large. Since these are characteristics of graphs of power law functions, then such behaviour suggests a law of this type. Exponential functions do not have these features, as the graphs show.



Power law models



Exponential models

It is not appropriate to determine the form of a model for a given set of data simply by using a process in which the data are linearised, a value of R^2 calculated for each, and then the model with the larger R^2 value chosen, as the R^2 values cannot be compared (refer to comments on R^2 in the section Notes on *R*-squared and *r* (pages 9 – 12) in the notes for achievement standard 90645).

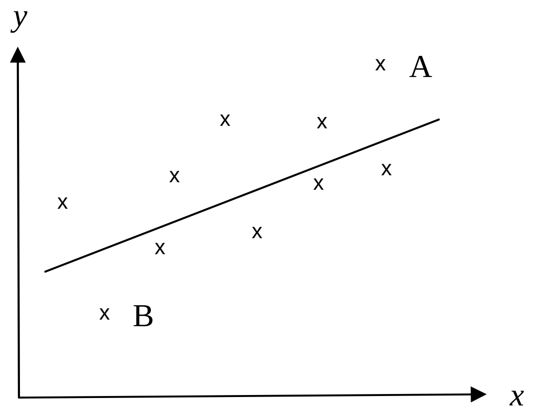
2 Determining the equation of the model

In cases where data that are modelled by a function of the form $y = Ae^{kx}$ are linearised and the data includes $x = 0$, the value of *A* must be the value of *y* when $x = 0$. For example, in using drawing pins to simulate radioactive decay, if there are 80 drawing pins initially, then *A* must be taken as 80 as this value is fixed; other observations arise randomly – if the experiment was repeated the observations could well be different. With Excel, the vertical intercept of the regression line can be set as 80 (under the menu option "Format trendline" and then "Options", there is an option to set the intercept).

Graphics calculators do not offer this possibility and so should not be used for the analysis in this case.

A procedural difficulty arises in cases of growth and decay where the data includes the value $x = 0$ but where attempts are being made to fit a power law (the difficulty is that the logarithm of 0 is not defined). Some teachers have adopted the practice of taking, say, $x = 0.1$, to circumvent the difficulty, but such practice is not appropriate - data cannot be amended to suit the method of analysis! The situation will not arise if due consideration is given to the likely form of model before undertaking any analysis.

If the gradient of the line of best fit to linearised data is found manually, it should be found using two points that lie on the line, with the values used in calculating the gradient being obtained from the graph. Using other points, specifically actual observations, may result in significant errors since the points chosen may not be close to the line, as shown in the following diagram if the points A and B are used.



EXCELLENCE

1 Relating the solution to the problem

One of the other main purposes of determining a model for data is to further understand the context from which the data arose. If an inappropriate model is chosen this may lead to incorrect conclusions about the context of the data being drawn.

An in-depth, well-developed discussion on selecting a model (a requirement for Merit) and how well the model fits the data could meet this aspect for Excellence. Selecting a model should be based primarily on an inspection of the fit of the model to the data, which can be supported by consideration of the value of R^2. Discussion could include how well the model fits the raw data over the entire range of the data.

It is not appropriate to test how well a model fits a set of observations by selecting individual data point(s) and comparing the observed value with the predicted value of the response variable for a particular value of the explanatory variable (this is like using an individual value of a variable to represent all values of the variable). Any experimental data most likely contains random variation, and the random variation in any individual value(s) chosen to test the model may be greater than that in the other values, so the point(s) chosen may be atypical. R^2 provides a measure of how well the model fits since it is calculated using all values of the response variable (in the same way that the mean usually provides the best single value that represents all values of a variable since it is calculated using all values of the variable).

Students should have an understanding that a different experiment with the same variables would probably lead to different data, and that the subsequent analysis would probably result in different values of the parameters (constants) being found. They need, therefore, to understand that the values of constants found are estimates of the values being sought.

In physical laws of the form $y = Ax^n$, the value of n is often integral or rational. Depending on the context, consideration could be given to the value of n in a power law of the form $y = Ax^n$. If the value of n as determined in the analysis is, say, -2.04, it may be appropriate to take $n = -2$ (so the law is taken to be of the form $y = \frac{A}{x^2}$). Similarly, if the value of n as determined is, say, 0.493, it may be appropriate to take $n = 0.5$ (so the law is taken to be of the form $y = A\sqrt{x}$).

2 Considering the nature of the underlying variables

Refer to comments for Merit. The requirements for this aspect of Excellence could be met by an in-depth discussion about the nature of the variables and how it led to the model that was selected.

3 Considering the possibility of more than one model

See the comments above about selecting models. The requirements for this aspect of Excellence could be met by an in-depth discussion about and comparison with one of more other models, including justification of the model(s) discussed.

Another model to select could be a piece-wise model. Its selection needs to be justified.

4 Relating the theory to the model

This could include in-depth discussion about why a model of the form $y = Ae^{kx}$ is appropriate for two variables x and y for which it can be assumed that $\frac{dy}{dx} \propto y$, including full discussion about the meaning of the statement $\frac{dy}{dx} \propto y$.

Another aspect that could be discussed under this heading is how the log transformation(s) linearise(s) the data, with an explanation of how the straight line on the graph of the transformed data then implies a particular model. This must be a well-developed argument, and must be in context - it is insufficient to just produce the theory behind the log transformation.

5 Discussing limitations of the model

See comments on this aspect in the section Limitations (pages 7 and 8) in the notes for achievement standard 90645, which are also relevant for this standard.